ONLINE APPENDIX FOR "NEURAL EVIDENCE OF REGRET AND ITS IMPLICATIONS FOR INVESTOR BEHAVIOR"

By

Cary Frydman and Colin Camerer

Appendix I. Experimental Instructions

Buying your stock

In this experiment you will be given 350 experimental dollars to invest in three different stocks. Your job is to choose when to buy and sell each stock, so that you earn the most money by the end of the experiment. Throughout the experiment, you will see the price of each stock changing (more detail below), and you will use this information to decide when to buy and sell. When you sell a stock, you receive an amount of cash equal to the price of the stock. When you buy a stock, you receive one unit of the stock, but you must give up an amount of cash equal to the current price of the stock.

The three stocks you can buy or sell are simply called Stock A, Stock B, and Stock C. To begin the experiment you *MUST* buy all three stocks, where each stock costs $100. Therefore, after you buy the three stocks, you will own one unit of each stock and have a total of $50 remaining. For the remainder of the experiment, you are only allowed to hold a maximum of 1 unit of each stock, and you cannot hold negative units (no short selling.) However, you can carry a *negative cash balance* by buying a stock for more money than you have, but any negative cash balances will be deducted from your final earnings.

Structure of the market

In the experiment, you will see two types of screens, a *price update screen* and an *action screen.* In the price update screen, one stock will be randomly selected and you will be told if the selected stock price has gone up or down, and by how much. Note that you will only see an update for *one* stock at a time. You will not be asked to do anything during this screen, you will simply see information about the change in price.

Following the price update screen, another stock will be randomly chosen (it may be the same one you just saw) and you will be asked to take an action. If you currently hold a unit of the stock, you will be asked if you would like to sell the stock at the current price. If you do not currently own a unit of the stock, you will be asked if you would like to buy a unit at the current price.

The experiment will start out with 9 consecutive price update screens, and then you will have the opportunity to buy or sell after each subsequent price update screen.

How the stock prices change

Each stock changes price according to the exact same rule. Each stock is either in a good state or in a bad state. In the good state, the stock goes up with 55% chance, and it goes down with 45% chance. In the bad state, the stock goes down with 55% chance and it goes up with 45% chance.

Once it is determined whether the price will go up or down, the *size* of the change is always random, and will either be $5, $10, or $15. For example, in the bad state, the stock will go down with 55% chance, and the amount it goes down by is $5, $10, or $15 with equal chance. Similarly, the good stock will go up with 55% chance, and the amount it goes up by will either be $5, $10, or $15.

The stocks will all randomly start in either the good state or bad state, and after each price update, there is a 20% chance the stock switches state.

Stock price changes

	Good state	*Bad state*
+	55%	45%
-	45%	55%

State changes

	Good state today	*Bad state today*
Good state tomorrow	80%	20%
Bad state tomorrow	20%	80%

Earnings and payout

You will play this market game TWO **SEPARATE** TIMES in the scanner. Each game will last approximately 15 minutes, and each game is independent from the previous one. This means when you start the second game, you will have to buy the three stocks at $100 again, and the stocks will start randomly in each state again.

Your earnings at the end of the experiment will be equal to the amount of cash you accrued over the two scanning sessions from buying and selling stocks, plus the current price of any stocks that you own.

Earnings=cash + price A(Hold A) + Price B*(Hold B) + Price C*(Hold C)*

Finally, your earnings will be converted using an exchange rate of 12:1. That means we divide your earnings by 12, and pay you this amount plus the $15 show up fee.

Button presses

During the Action screens, you will either be given the option to "Buy?" or "Sell?" depending on whether you hold the stock or not. The LEFT (blue) button indicates "YES". And the RIGHT (yellow) button indicates "NO." You have three seconds to enter your response, otherwise the computer will randomly select a response for you.

Appendix II. fMRI Data Collection and Analysis

In this section, we describe how the fMRI measures of neural activity were collected and analyzed. The goal of this section, which is primarily taken from Frydman et al. (2014), is to provide enough information to serve as a brief primer on the subject for readers who are unfamiliar with fMRI. For a more detailed discussion, see (Huettel, Song and McCarthy (2004); Ashby (2011); Poldrack, Mumford and Nichols (2011)).

A. fMRI Data Collection and Measurement

We collected measures of neural activity over the entire brain using BOLD-fMRI, which stands for blood-oxygenated level dependent functional magnetic resonance imaging. BOLD-fMRI measures changes in local magnetic fields that result from the local inflows of oxygenated hemoglobin and outflows of de-oxygenated hemoglobin that occur when neurons fire. In particular, fMRI provides measures of the BOLD response in small "neighborhoods" of brain tissue called *voxels*, and is thought to measure the sum of the total amount of neuronal firing into that voxel and the total amount of neuronal firing within the voxel.[1]

One important complication is that the hemoglobin responses measured by BOLD-fMRI are slower than the associated neuronal responses. Specifically, although the bulk of the neuronal response takes place quickly, BOLD measurements are affected for up to 24 seconds thereafter. Panel A of Figure A1 provides a more detailed illustration of the nature of the BOLD response. It depicts the path of the BOLD signal in response to one (arbitrary) unit of neural activity of infinitesimal duration at time zero. The function plotted here is called the canonical hemodynamic response function (HRF). It is denoted by $h(\tau)$, where τ is the amount of time elapsed since the

[1] The neural activity measured by fMRI in a 1 mm^3 cube (about the size of a grain of salt) represents the joint activity of between 5,000 to 40,000 neurons, depending on the area of the brain.

neural activity impulse, and has been shown to approximate well the pattern of BOLD responses for most subjects, brain areas, and tasks.

Fortunately, there is a standard way of dealing with the complication described in the previous paragraph. In particular, the BOLD response has been shown to combine linearly across multiple sources of neural activity (Boynton et al. (1996)). This property, along with knowledge of the specific functional form of the HRF, allows us to construct a mapping from predicted neural activity to predicted BOLD responses. Specifically, if the predicted level of neural activity at any particular time is given by $a(t)$, then the level of BOLD activity at any instant t is well approximated by

$$b(t)=\int_0^\infty h(u)a(t-u)\,du, \qquad \text{(A1)}$$

which is the convolution between the HRF and the neural inputs. This integral has a straightforward interpretation: it is a lagged sum of all the BOLD responses triggered by previous neural activity. Panel B of Figure A1 illustrates the connection between neural activity and BOLD responses; it depicts a hypothetical path of neural activity (the solid line), together with the associated BOLD response (the dashed line).

During our experiment, we acquire two types of MRI data in a 3.0 Siemens Tesla Trio MRI scanner with an eight-channel phased array coil. First, we acquire BOLD-fMRI data while the subjects perform the experimental task. We use a voxel size of 3 mm^3, and collect these data for the entire brain (~ 100,000 voxels) every 2.75 seconds.[2] We also acquire high-resolution

[2] More precisely, we acquire gradient echo T2*-weighted echoplanar (EPI) images with BOLD contrast. To optimize functional sensitivity in the orbitofrontal cortex (OFC), a key region of interest, we acquire the images in an oblique orientation of 30° to the anterior commissure–posterior commissure line [Deichmann, 2003 #16. Each volume of images has 45 axial slices. A total of 692 volumes were collected over two sessions. The imaging parameters are as follows: echo time, 30 ms; field of view, 192 mm; in-plane resolution and slice thickness, 3 mm; repetition time, 2.75 s.

anatomical scans that we use mainly for realigning the brains across subjects and for localizing the brain activity identified by our analyses.[3]

B. fMRI Data Pre-processing

Before the BOLD data can be analyzed to test our hypotheses, they have to be converted into a usable format. This requires the following steps, which are fairly standard – see Huettel, Song, and McCarthy (2004), Ashby (2011), and Poldrack, Mumford, and Nichols (2011) – and which are implemented by way of a specialized but commonly-used software package called SPM5 (Wellcome Department of Imaging Neuroscience, Institute of Neurology, London, UK).

First, we correct for slice acquisition time within each voxel. This is necessary because the scanner does not collect data on all brain voxels simultaneously. This simple step, which involves a nonlinear interpolation, realigns the data across all voxels.

Second, we correct for head motion to ensure that the time series of BOLD measurements recorded at a specific spatial location within the scanner is always associated with the same brain location throughout the experiment.[4]

Third, we realign the BOLD responses for each individual into a common neuroanatomical frame (the standard Montreal Neurological Institute EPI template). This step, called spatial normalization, is necessary because brains come in different shapes and sizes; as a result, a given spatial location maps to different brain regions in different subjects. Spatial

[3] More precisely, we acquire high-resolution T1-weighted structural scans (1 x 1 x 1 mm) for each subject. These are coregistered with their mean EPI images and averaged across subjects to permit anatomical localization of the functional activations at the group level.

[4] BOLD measurements were corrected for head motion by aligning them to the first full brain scan and normalizing to the Montreal Neurological Institute's EPI template. This entails estimating a six-parameter model of head motion for each volume (three parameters for center movement, and three parameters for rotation), and then removing the effect of the motion using these parameters. For details, see (Friston et al. (1996)).

normalization involves a nonlinear reshaping of the brain to maximize the match with a target template. Although the transformed data are not perfectly aligned across subjects due to remaining neuroanatomical heterogeneity, the process is sufficiently accurate for the purposes of most studies. Furthermore, any imperfections in the realignment process introduce noise that reduces our ability to detect neural activity of interest.

Fourth, we also spatially smooth the BOLD data for each subject by making BOLD responses for each voxel a weighted sum of the responses in neighboring voxels, where the weights decrease with distance.[5] This step ensures that the error structure of the data conforms to the normality assumptions on the error structure of the regression models that we will use to test our hypotheses (Huettel et al. (2004); Poldrack et al. (2011)).

Finally, we remove low-frequency signals that are unlikely to be associated with neuronal responses to individual trials.[6] An example of such a signal is the effect of a continuous head movement over the course of the experiment that is not fully removed by the second correction step described above.

C. fMRI Main Data Analyses

The key goals of our analysis are to test if the region of the vSt that has been repeatedly shown to encode prediction errors is consistent with Predictions 2 and 3.To do this, we run statistical tests to see if there are areas within these regions of the brain, given by collections of spatially contiguous voxels called *clusters*, where the BOLD response reflects neural activity that implements the computations of interest (e.g., realization utility computations). This is complicated by the fact that, since every voxel contains thousands of neurons, the BOLD

[5] Spatial smoothing was performed using an 8 mm full-width half-maximum Gaussian kernel. Essentially, this step entails replacing every measurement at every voxel with a weighted sum of the measurements in a neighborhood centered on the voxel, using weights that are given by the Gaussian kernel.

[6] Specifically, we applied a high-pass temporal filter to the BOLD data with a cut-off of 128 seconds.

responses in a voxel can be driven by multiple signals. Fortunately, the linear properties of the BOLD signal allow the neural signals of interest to be identified using standard linear regression methods.

The general statistical procedure is straightforward, and will be familiar to most economists. The analysis begins by specifying two types of variables that might affect the BOLD response: target computations and additional controls. The target computations reflect the signals we are looking for (e.g., a realization utility signal at the time of selling a stock). They are specified by a time series $s_i(t)$ describing each signal of interest. For each of these signals, let $S_i(t)$ denote the time series that results from convolving the signal $s_i(t)$ with the HRF, as described above. The additional controls, denoted by $c_j(t)$, are other variables that might affect the BOLD time series (e.g., residual head movement or time trends). These are introduced to further clean up the noise in the BOLD signal, but are not explicitly used in any of our tests. The control variables are not convolved with the HRF because, while they affect the measured BOLD responses, they do not reflect neural activity which triggers a hemodynamic response.[7]

The linearity of the BOLD signal implies that the level of BOLD activity $b^v(t)$ in any voxel v at time t should be given by

$$b^v(t) = \text{constant} + \sum_i \beta_i^v S_i(t) + \sum_j \alpha_j^v c_j(t) + \varepsilon(t), \qquad \text{(A2)}$$

where $\varepsilon(t)$ denotes AR(1) noise. This model is estimated independently in each of the voxels that fall within the relevant region of interest (the vSt). Our hypotheses can then be restated as tests about the coefficients of this regression model: signal i is said to be associated with activity in voxel v only if β_i^v is significantly different from zero.

[7] For example, linear trends are often included as controls because the scanner heats up with continuous operation, inducing a linear change in the measured BOLD responses.

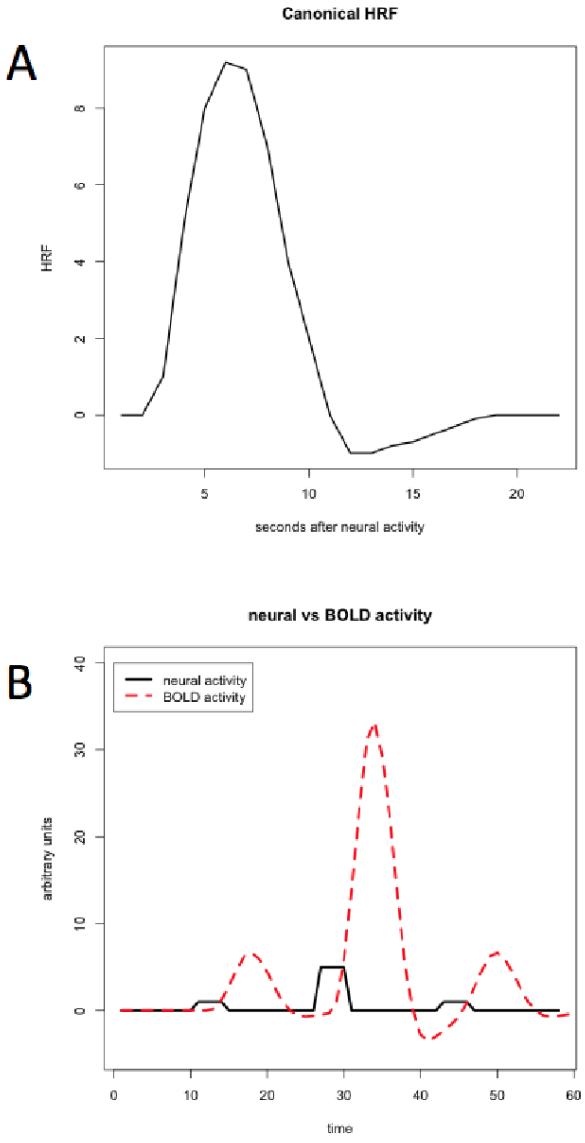
Two additional considerations apply to most fMRI studies, including this one. First, we are interested in testing hypotheses about the distribution of the signal coefficients in the population of subjects, not hypotheses about individual subject coefficients. This would normally require estimating a mixed effects version of the linear model specified above, which, given the size of a typical fMRI dataset, would be computationally intensive. Fortunately, there is a shortcut that provides a good approximation to the full mixed effects analysis (Penny et al. (2006)). It involves estimating the parameters separately for each individual subject, averaging them across subjects, and then performing *t*-tests. This is the approach we follow here.

Second, since our tests are carried out in each of the voxels in the relevant regions of interest (68 for the vSt), there is a concern about false-positives. To address this problem, we correct for multiple comparisons within the relevant region of interest, a procedure known in the fMRI literature as a small volume correction (SVC). We report results as significant if they pass SVC correction at a level of $p<0.05$.[8]

As noted earlier, we conduct our tests in an area of the vSt that has been linked to the computation of prediction errors. Specifically, we construct a sphere with a 15 mm radius around the coordinates (MNI-space, $x = -15$, $y = 6$, $z = -12$) that were found to exhibit peak correlation with prediction errors in (Lin, Adolphs and Rangel (2012)), and then intersect this sphere with an anatomical mask of the vSt.

[8] Specifically, we report results as significant if voxels within the pre-specified region of interest pass $p<0.005$ uncorrected with a 20-voxel extent threshold and if they pass SVC with a family-wise error rate of less than 0.05.

Figure A1. **BOLD measurements of neural activity.** Panel A: Because fMRI measures the blood-oxygenated level dependent (BOLD) response, and not neural activity itself, we need a mapping from neural activity to BOLD response to make inferences about changes in neural activity. This mapping is known as the canonical hemodynamic response function and is shown here as the result of one unit of instantaneous neural activity at time 0. Panel B: This figure shows the BOLD response that results from three sequential sources of neural activity. The BOLD response combines linearly across multiple sources of neural activity.



REFERENCES

1. S. Huettel, A. Song, G. McCarthy, *Functional Magnetic Resonance Imaging*. (Sinauer Associates, 2004).
2. F. G. Ashby, *Statistical Analysis of fMRI Data*. (The MIT Press, 2011).
3. R. A. Poldrack, J. Mumford, T. Nichols, *Handbook of Functional MRI Data Analysis*. (Cambridge University Press, 2011).
4. G. M. Boynton, S. A. Engel, G. H. Glover, D. J. Heeger, Linear Systems Analysis of Functional Magnetic Resonance Imaging in Human V1. *The Journal of Neuroscience* **16**, 4207-4221 (1996).
5. K. J. Friston, A. Holmes, J.-B. Polin, C. J. Price, C. D. Frith, Detecting activations in PET and fMRI: Levels of inference and power. *Neuroimage* **40**, 223-235 (1996).
6. W. Penny, K. Friston, J. Ashburner, S. Kiebel, T. Nichols, *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. (Academic Press, 2006).
7. A. Lin, R. Adolphs, A. Rangel, Social and monetary reward learning engage overlapping neural substrates. *Social Cognitive and Affective Neuroscience* **7**, 274-281 (2012).